Appendix

Algorithm 1 Reconstructing survival data (adapted from Guyot et al. [2012])

Require: The data extracted from published survival curves.

S_k: survival percentages as extracted from y axis, $k = 1, \ldots, K$, where K is the total number of extracted data points

t_k: time from randomization as extracted from x axis

$nrisk_i$: number of patients at risk at time $trisk_i, i = 1, \ldots, T$, where T is the number of intervals where the number of patients at risk is reported

$trisk_i$: time reported at the risk table

Ensure: $S_{k+1} \leq S_k$ for all k to meet the monotonicity constraint.

Set $\text{lower}_i = \min\{k : t_k \geq trisk_i\}$ and $\text{upper}_i = \max\{k : t_k \leq trisk_{i+1}\}$.

if $i < T - 1$ and $T > 1$ **then**

Step 1. Calculate $\widehat{nc}_i$, the number of censored at time $[trisk_i, trisk_{i+1}]$, by

$$\widehat{nc}_i = S_{\text{lower}_{i+1}} / S_{\text{lower}_i} \times nrisk_i - nrisk_{i+1}$$

Step 2. Distribute $\widehat{nc}_i$ evenly within $[trisk_i, trisk_{i+1}]$. The censored time is then

$$\widehat{ctime}_c = t_{\text{lower}_i} + c \times (t_{\text{lower}_{i+1}} - t_{\text{lower}_i})/(\widehat{nc}_i + 1)$$

where $c = 1, \ldots, \widehat{nc}_i$. We can then calculate the number of censored events, $\widehat{nc}_k$, in extracted intervals $[t_k, t_{k+1}]$, which is within $[trisk_i, trisk_{i+1}]$.

Step 3. Calculate the number of events at t_k as

$$\widehat{nd}_k = \widehat{n}_k \times \left(1 - S_k / \widehat{S}^{KM}_{\text{last}(k)}\right)$$

$\widehat{n}_k$ is the estimated number at risk at time t_k. $\widehat{S}^{KM}_{\text{last}(k)}$ is the estimated survival probability at time $t_{\text{last}(k)}$ with

$$\text{last}(k) = \begin{cases} 1 & \text{if } k = 1 \\ k' & \text{otherwise} \end{cases}$$

Note that $t_{k'} \leq t_k$, k' is such that the latest event occurs at $t_{k'}$, and there are no events in $(t_{k'}, t_k)$. The estimated number of patients at risk at time t_{k+1} is then $\widehat{n}_{k+1} = \widehat{n}_k - \widehat{nd}_k - \widehat{nc}_k$, where $k \in [\text{lower}_i, \text{upper}_i]$. Thus, $\widehat{nrisk}_{i+1} = \widehat{n}_{\text{upper}_i + 1}$.

Step 4. Set $\Delta_t = \widehat{nrisk}_{i+1} - nrisk_{i+1}$.

if $\Delta_t \neq 0$ **then**

Adjust the estimated number of censored in time interval $[trisk_i, trisk_{i+1}]$ by setting

$$\widehat{nc}_i = \widehat{nc}_i + \left(\widehat{nrisk}_{i+1} - nrisk_{i+1}\right)$$

We then repeat steps 1–4 until $\widehat{nrisk}_{i+1} = nrisk_{i+1}$.

end if

Step 5. Repeat steps 1–4 until $i + 1 = T$.

end if

if $i = T$ or $i = 1$ and $T = 1$ **then**

Step 6. Approximate $\widehat{nc}_T$ within interval $[trisk_{T-1}, trisk_T]$ by setting

$$\widehat{nc}_T = \min\left(\frac{t_{\text{upper}_T} - t_{\text{lower}_T}}{t_{\text{upper}_{T-1}} - t_{\text{lower}_1}} \times \sum_{i=1}^{T-1} \widehat{nc}_i;\ nrisk_T\right)$$

We then run steps 2–3 for the last interval $[trisk_{T-1}, trisk_T]$.

end if

if the total number of events, D, is not given **then**

Stop the algorithm.

end if

if the total number of events, D, is given **then**

Step 7. Compute $\sum_{k=1}^{\text{upper}_{T-1}} \widehat{nd}_k$.

if $\sum_{k=1}^{\text{upper}_{T-1}} \widehat{nd}_k \geq D$ **then**

Stop the algorithm.

end if

if $\sum_{k=1}^{\text{upper}_{T-1}} \widehat{nd}_k < D$ **then**

Step 8. Adjust the number of censored, $\widehat{nc}_T$, by setting

$$\widehat{nc}_T = \widehat{nc}_T + \left(\sum_{k=1}^{\text{upper}_T} \widehat{nd}_k - D\right)$$

Repeat steps 2–3 and steps 7–8 for the last interval.

end if

end if